



کاربرد XGBoost برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران

شهلا ابراهیمی

دانشجوی دکتری حسابداری دانشگاه شیراز، ایران

دکتر محمد نمازی^۱

استاد ممتاز حسابداری دانشگاه شیراز، ایران

(تاریخ دریافت: ۱۷ فروردین ۱۳۹۹؛ تاریخ پذیرش: ۷ تیر ۱۳۹۹)

هدف این مقاله، پیش‌بینی درماندگی مالی بالقوه شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار است. بدین منظور، دامنه گسترده‌ای از ویژگی‌ها از جمله متغیرهای حسابداری تعهدی، حسابداری نقدی، بازار سهام، مکانیسم‌های حاکمیت شرکتی و شاخص‌های اقتصاد کلان برای پیش‌بینی درماندگی مالی شرکت‌های نمونه شناسایی شده‌اند. نمونه نهایی شامل ۴۲۱ شرکت و در نتیجه، ۳۶۷۰ شرکت-سال مشاهده است. سپس، داده آماده شده با استفاده از نسبت ۷۰ به ۳۰ به مجموعه داده آموزشی و آزمایشی تفکیک شد. در این پژوهش، تکنیک‌های پیش پردازش داده یادگیری ماشین نظیر استانداردسازی نمره Z، وان-هات انکدینگ، اعتبارسنجی متقابل K لایه طبقه‌ای، همراه با مهندسی ویژگی برای بهبود عملکرد طبقه‌بندی کننده بکار گرفته شدند. روش اعتبارسنجی متقابل K لایه طبقه‌ای با $(K=5)$ برای برآورد عملکرد پیش‌بینی مدل طی مرحله آموزش استفاده شد. طی مرحله آموزش، میزان‌سازی ابرپارامتر مدل با استفاده از جستجوی شبکه‌ای انجام شد. افزون بر این، تکنیک SMOTE همراه با معیار مختص مسائل نامتوازن یعنی نمره F1 برای غلبه بر مسأله نامتوازی افراطی کلاس‌ها استفاده شده است. بر اساس نتایج تجربی، مدل XGBoost به نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۹۰٪، ۹۰٪، ۱۰۰٪ و ۸۲٪ بر روی مجموعه آموزشی دست یافت. سرانجام، مدل پیشنهادی بر روی مجموعه آزمایشی کنار گذاشته شده آزمون شد که به نمره F1، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۲٪، ۵۲٪، ۷۳٪ و ۴۱٪ بر روی مجموعه آزمایشی منجر شد. این اطلاعات، ابزار قدرتمندی برای پیش‌بینی درماندگی مالی شرکت‌ها فراهم می‌کنند.

واژه‌های کلیدی: پیش‌بینی درماندگی مالی، XGBoost، یادگیری ماشین، داده‌کاوی، بورس اوراق بهادار تهران.

¹ mnamazi@rose.shirazu.ac.ir

مقدمه

از آنجا که امروزه محیط کسب و کار به میزان قابل توجهی نامطمئن و رقابتی شده است، ورشکستگی برای هر سازمانی می‌تواند اتفاق بیفتد [13]. ورشکستگی یک مسأله بسیار با اهمیت جهانی و دارای هزینه‌های اجتماعی زیاد است [13]. تشخیص درماندگی تجاری و هشدار زودهنگام درباره وقوع بحران مالی تنها برای تحلیل‌گران و دست‌اندرکاران حرفه در ایالات متحده حائز اهمیت نیست. در واقع، کشورهای سراسر جهان، حتی ملت‌های غیر سرمایه‌داری نیز نگران ارزیابی عملکرد شرکت‌ها هستند. کشورهای در حال توسعه و اقتصادهای کوچک‌تر و همچنین، ملت‌های صنعتی بزرگ‌تر جهان به شدت خواهان اجتناب از بحران مالی در بخش‌های خصوصی و دولتی هستند. ملت‌های کوچک‌تر نسبت به بحران‌های مالی ناشی از درماندگی واحدهای تجاری به شدت آسیب‌پذیر هستند. از این رو، برخی سیاستمداران ملت‌های کوچک‌تر به طور خاص نگران سراسیمگی‌ها و هراس‌زدگی‌های مالی ناشی از درماندگی تک تک واحدهای تجاری هستند [4, 135, 2020].

این واقعیت موجب می‌شود که توسعه و استفاده از مدل‌های پیش‌بینی ورشکستگی صحیح، ابزار بازدارنده ارزشمندی برای بستانکاران، مدیران، مالکان و سهامداران باشد. اگرچه هیچ روش تضمین شده‌ای برای پیش‌بینی ورشکستگی شرکت وجود ندارد اما مدل‌های بسیاری وجود دارد که موفقیت آن‌ها در تجزیه و تحلیل ورشکستگی اثبات شده است [132020]. از این رو، هدف این پژوهش، درصدد بکارگیری یادگیری ماشین جهت ارائه هشدار زودهنگام درماندگی مالی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از داده‌های مبتنی بر ترازنامه و صورت سود (زیان)، صورت گردش وجوه نقد، بازار سهام و نقدشوندگی، حاکمیت شرکتی و متغیرهای کلان اقتصادی به ذینفعان است تا آن‌ها بتوانند اقدامات مقتضی جهت پیشگیری از درماندگی مالی و همچنین، کاهش هزینه‌ها و زیان‌های ناشی از درماندگی مالی انجام دهند. افزون بر این، با توجه به تنوع و گستردگی متغیرهای پیش‌بینی کننده درماندگی مالی، تعیین اهمیت این متغیرها و انتخاب مربوط‌ترین متغیرهای حاوی اطلاعات مفید نیز از جمله اهداف این پژوهش است. اهمیت این پژوهش این است که به معرفی یک الگوی قوی و با دقت بالا جهت پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از فن یادگیری ماشین می‌پردازد و می‌تواند مورد استفاده بورس، شرکت‌ها، سرمایه‌گذاران، تحلیل‌گران مالی و پژوهشگران حسابداری و مالی قرار گیرد. در این پژوهش ابتدا مروری بر ادبیات داخلی و خارجی و مبانی نظری انجام خواهد شد. سپس روش شناسی پژوهش، نحوه گردآوری اطلاعات، یافته‌ها و نتیجه‌گیری و پیشنهادها ارائه می‌شود.

مروری بر پیشینه پژوهش

پیشینه پژوهش داخلی

اعتمادی، انواری رستمی و فرج زاده دهکردی (۲۰۰۹) با بکارگیری برنامه‌نویسی ژنتیک و تحلیل تشخیصی چندمتغیره به ارائه مدلی برای پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران پرداختند. نمونه مورد بررسی آن‌ها شامل ۷۲ شرکت ورشکسته و ۷۲ شرکت غیرورشکسته طی

دوره زمانی ۱۹۹۸-۲۰۰۸ بود. مدل برنامه‌نویسی ژنتیک آن‌ها به نرخ صحتی برابر با ۹۴٪ و ۹۰٪ به ترتیب در خصوص نمونه‌های آموزشی و آزمایشی دست یافت در حالی که نرخ صحت مدل تحلیل تشخیصی چندمتغیره در نمونه‌های آموزشی و آزمایشی به ترتیب برابر با ۷۷٪ و ۷۳٪ بود. آزمون مک‌نمار نیز نشان داد که عملکرد رویکرد برنامه‌نویسی ژنتیک در خصوص مسأله ورشکستگی شرکت، بهتر از عملکرد تحلیل تشخیصی چندمتغیره است.

احمدپور و میرزایی اسرمی (۱۳۹۲) به مقایسه مدل تحلیل تمایزی چندگانه با مدل شبکه‌های عصبی برای پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بازار بورس اوراق بهادار تهران پرداختند. نتایج حاصل از بکارگیری مدل‌های مذکور بر روی نمونه‌ای شامل ۵۴ شرکت ورشکسته و ۶۴ شرکت غیرورشکسته طی دوره زمانی ۱۳۷۹-۱۳۸۹ نشان‌دهنده این است که توان پیش‌بینی مدل شبکه‌های عصبی برتر از مدل تحلیل تمایزی است.

پیشینه پژوهش خارجی

باربوزا، کیمورا و آلتمن^۱ (۲۰۱۷) به مقایسه عملکرد مدل‌های یادگیری ماشین (ماشین بردار پشتیبان، بگینگ، بوستینگ، جنگل تصادفی) با تحلیل تشخیصی، رگرسیون لجستیک و شبکه‌های عصبی برای پیش‌بینی ورشکستگی شرکت‌های آمریکای شمالی طی دوره زمانی ۱۹۸۵-۲۰۱۳ پرداختند. آن‌ها دریافتند که به طور میانگین، صحت مدل‌های یادگیری ماشین تقریباً ۱۰٪ بیشتر از مدل‌های سنتی است. همچنین، افزودن متغیرهای اضافی نظیر حاشیه عملیاتی، تغییر در بازده حقوق صاحبان سهام، تغییر در نسبت قیمت به ارزش دفتری و معیارهای سنجش رشد دارایی‌ها، فروش و تعداد کارکنان به متغیرهای مدل Z آلتمن موجب بهبود صحت پیش‌بینی نمونه آزمایشی می‌شود.

چارالامبکیس و گارت^۲ (۲۰۱۸) به بررسی احتمال درماندگی مالی ۳۰۸۸۶ شرکت سهامی خاص یونان طی دوره زمانی ۲۰۰۳ تا ۲۰۱۱، در مجموع ۱۸۸۰۶۵ مشاهده سالانه با استفاده از مدل لوجیت خطر گسسته (چند دوره‌ای) پرداختند. یافته‌های آن‌ها بر روی ۱۷۷۰ شرکت ورشکسته با ۵۹۵۷ مشاهده شرکت-سال و ۲۹۱۱۶ شرکت غیرورشکسته با ۱۸۲۱۰۸ مشاهده شرکت-سال بیانگر این است که سودآوری، اهرم، نسبت سود انباشته به کل دارایی‌ها، اندازه، نسبت نقدینگی، متغیر دامی صادرات، تمایل به پرداخت سود سهام و نرخ رشد تولید ناخالص داخلی واقعی، پیش‌بینی کننده‌های قوی برای احتمال درماندگی مالی شرکت‌ها هستند. افزون بر این، عملکرد مدل حاوی این متغیرها از نظر صحت طبقه‌بندی شرکت‌ها دارای بیشترین صحت است و با افزایش افق پیش‌بینی نیز توانایی پیش‌بینی مدل حفظ می‌شود.

ادبیات و چارچوب نظری

در همه انواع بحران‌های مالی، سرانجام عملکرد ضعیف شرکت در صورت‌های مالی منعکس می‌شود و با استفاده از این صورت‌ها می‌توان نسبت‌های مالی مفیدی را برای ارزیابی ورشکستگی محاسبه کرد

¹ Barboza, Kimura and Altman

² Charalambakis and Garrett

[1336]. از این رو، تقریباً تمامی مدل‌های پیش‌بینی بحران مالی به پیش‌بینی ورشکستگی یا بحران مالی بر حسب صورت‌های مالی می‌پردازند. در واقع، همیشه اطلاعات بسیار زیادی در صورت‌های مالی شرکت افشا می‌شود. در نتیجه، این امر که اطلاعات برگرفته از صورت‌های مالی شرکت به صورت مستقیم یا غیرمستقیم، شاخص‌های خوبی برای پیش‌بینی بحران مالی شرکت هستند، بسیار با اهمیت است. بسیاری از پژوهش‌گران حسابداری و مالی به بررسی نسبت‌ها یا متغیرهای مختلفی پرداخته‌اند که برای پیش‌بینی بحران مالی شرکت سودمند هستند [1338].

درماندگی مالی به معنای ناتوانی شرکت به ایفای تعهدات مالی در تاریخ سررسید است. بر اساس این تعریف، شرکت در صورتی دچار درماندگی مالی می‌شود که نتواند وجه نقد کافی برای تأمین نیازهای خود نظیر بازپرداخت به اعتباردهندگان و دارندگان اوراق بدهی ایجاد کند. از این رو، انتظار می‌رود که اطلاعات جریان وجوه نقد برای پیش‌بینی درماندگی مالی مفید باشد [17171333].

از آنجا که گزارش مالی شرکت می‌تواند در معرض دست‌کاری مدیریت قرار گیرد، این گزارش از طریق مدیریت سود دستخوش حساب آرای می‌شود. دعوای حقوقی اخیر علیه ورلداکام درباره ارائه بیشتر از اندازه سود به میزان ۹ میلیارد دلار از سال ۱۹۹۹ تا فصل نخست سال ۲۰۰۲، نمونه‌ای از دست‌کاری ارقام حسابداری توسط مدیریت [1316] و بیانگر مشکلات عدم رعایت مقررات حسابداری و فقدان سیستم نظارتی خوب است. افزون بر این، شکست انرون در سال ۲۰۰۱ نیز به دلیل مکانیسم‌های حاکمیت شرکتی ضعیف بود [1334]. بنابراین، متغیرهای حسابداری به تنهایی ممکن است به گونه کافی درماندگی مالی شرکت را توضیح ندهند. از این رو، ساختار مالکیت و حاکمیت شرکتی نیز متغیرهای مهمی هستند که به توضیح درماندگی مالی شرکت کمک می‌کنند [1316].

افزون بر این، گنجاندن متغیرهای مبتنی بر بازار و نقدشوندگی سهام در مدل‌های پیش‌بینی درماندگی مالی به چند دلیل جذاب و لازم است. نخست، قیمت‌های بازار بیانگر ترکیب غنی و جامعی از اطلاعاتی است که داده‌های صورت‌های مالی تنها یک زیرمجموعه از آن است. دوم، متغیرهای مبتنی بر بازار را می‌توان در محدوده زمانی کوچک‌تری اندازه‌گیری نمود. در حالی که صورت‌های مالی در بهترین حالت بر مبنای فصلی در دسترس هستند و پژوهش‌های پیشین عمدتاً از داده‌های سالانه استفاده کرده‌اند، متغیرهای مبتنی بر بازار می‌توانند از دسترسی به قیمت‌های روزانه بهره‌مند شوند. سوم، متغیرهای مبتنی بر بازار می‌توانند برآوردهایی از نوسان را فراهم کنند [137].

همچنین، وخامت و پسرفت محیط اقتصادی موجب افزایش ریسک سیستماتیک همه شرکت‌ها می‌شود. اما، شرکت‌های حساس‌تر نسبت به ناپایداری اقتصاد کلان در مقابل بحران مالی آسیب‌پذیرتر هستند. بنابراین، گنجاندن چندین معیار سنجش حساسیت شرکت‌ها نسبت به وضعیت اقتصاد کلان در مدل‌های پیش‌بینی درماندگی مالی نیز سودمند است [1322].

پرسش‌های پژوهش

با توجه به پیشینه و مبانی نظری، این پژوهش به بررسی این می‌پردازد که:

۱- آیا متغیرهای حسابداری، جریان وجوه نقد، بازار و نقدشوندگی، حاکمیت شرکتی و اقتصاد کلان می‌توانند بین شرکت‌های درمانده مالی و شرکت‌های موفق مالی تمایز ایجاد کنند؟ به بیان دقیق‌تر، آیا این متغیرها، پیش‌بینی کننده درماندگی مالی هستند؟

۲- کدام یک از متغیرهای حسابداری، جریان وجوه نقد، بازار و نقدشوندگی، حاکمیت شرکتی و اقتصاد کلان، برای استفاده‌کنندگان در پیش‌بینی درماندگی مالی مفید هستند؟

روش شناسی پژوهش

نوع مطالعه و روش بررسی پرسش‌ها

از آنجا که این پژوهش، مبتنی بر داده است و نتایج آن از طریق مشاهده قابل تأیید است، در قالب پژوهش تجربی قرار می‌گیرد [1319]. افزون بر این، با توجه به اینکه درصدد پاسخگویی به سؤال‌های دنیای واقعی است، کاربردی [1318] و به دلیل بکارگیری داده‌های کمی، در زمره پژوهش کمی محض جای می‌گیرد [1318]. همچنین، این پژوهش، پژوهش غیرآزمایشی طولی (آینده‌نگر) پیش‌بینی کننده است. غیرآزمایشی از این جهت که هیچ گونه دست‌کاری یا تخصیص تصادفی وجود ندارد [139]. به دلیل جمع‌آوری داده‌های هر شرکت در چندین دوره زمانی، نیز پژوهش طولی از نوع پنل است [1318]. پیش‌بینی کننده به این دلیل که درصدد پیش‌بینی متغیر هدف بر مبنای چندین متغیر پیش‌بینی کننده است [1318].

متغیرهای پژوهش

در این پژوهش، از دو دسته متغیر استفاده می‌شود:

متغیر مستقل: رویکرد معمول برای پیش‌بینی ورشکستگی این است که برای شناسایی مجموعه بزرگی از متغیرهای مستقل مالی یا غیرمالی بالقوه به مرور متون پرداخت و سپس، از طریق ترکیب تحلیل‌های ریاضیاتی و قضاوتی به مجموعه کوچک‌تری از متغیرهای مستقل دست یافت که ورشکستگی را پیش‌بینی می‌کنند [1326]. بنابراین، پس از ۳۰ سال پژوهش درباره این موضوع، هیچ مدل پذیرفته شده کلی برای پیش‌بینی ورشکستگی وجود ندارد که اساس آن مبتنی بر تعیین روابط علی عوامل اقتصادی زیربنایی باشد [1325، 2626]. در این پژوهش، سه معیار جهت انتخاب نسبت‌های مالی برای پیش‌بینی بحران مالی استفاده شده است:

۱- بیش از ۸۵۲ متغیر یا نسبت مختلف از ۱۲۸ مقاله استخراج گردید. انتخاب تمامی این متغیرها برای ایجاد مدل غیرممکن است. از این رو، نسبت‌ها یا عواملی که بیشتر از سه دفعه در ۱۲۸ مقاله مورد استفاده قرار گرفته‌اند، انتخاب خواهند شد.

۲- کشورهای مختلف ممکن است سیاست‌های حسابداری مختلفی داشته باشند و شرکت‌های مختلف متعلق به صنایع مختلف ممکن است انواع مختلفی از صورت‌های مالی داشته باشند. از این رو، معمولاً تمامی ۸۵۲ متغیر برای همه شرکت‌ها در همه پایگاه داده‌ها در دسترس نیست [133838].

۳- تصمیم پژوهشگر بر مبنای تجربه قبلی خویش در پژوهش‌های پیشین یا بر مبنای آزمایش‌های مقدماتی با در نظر گرفتن این سه معیار، ۱۲۴ متغیر پیش‌بین به شرح مندرج در جدول ۱ انتخاب شد.

جدول ۱. متغیرهای پیش‌بین (ورودی‌ها یا ویژگی‌ها)

متغیر	طبقه
۱- نسبت جاری = $\frac{\text{دارایی‌های جاری}}{\text{بدهی‌های جاری}}$ ، ۲- بازده دارایی‌ها = $\frac{\text{سود خالص}}{\text{کل دارایی‌ها}}$ ، ۳- نسبت بدهی = $\frac{\text{کل بدهی‌ها}}{\text{کل دارایی‌ها}}$	نسبت‌های مالی مبتنی بر ترازنامه و صورت سود (زبان)
۴- گردش کل دارایی‌ها = $\frac{\text{فروش}}{\text{کل دارایی‌ها}}$ ، ۵- بازده عملیاتی دارایی‌ها = $\frac{\text{سود عملیاتی}}{\text{کل دارایی‌ها}}$ ، ۶- سرمایه در گردش به کل دارایی‌ها = $\frac{\text{سرمایه در گردش}}{\text{کل دارایی‌ها}}$ ، ۷- نسبت سریع = $\frac{\text{دارایی‌های سریع}}{\text{بدهی‌های جاری}}$ ، ۸- حاشیه سود خالص = $\frac{\text{سود خالص}}{\text{فروش}}$ ، ۹- بازده حقوق صاحبان سهام = $\frac{\text{سود خالص}}{\text{حقوق صاحبان سهام}}$ ، ۱۰- سود انباشته به کل دارایی‌ها = $\frac{\text{سود انباشته}}{\text{کل دارایی‌ها}}$ ، ۱۱- دارایی‌های جاری به کل دارایی‌ها = $\frac{\text{دارایی‌های جاری}}{\text{کل دارایی‌ها}}$ ، ۱۲- نسبت پوشش بهره = $\frac{\text{سود عملیاتی}}{\text{هزینه بهره}}$ ، ۱۳- حقوق صاحبان سهام به کل دارایی‌ها = $\frac{\text{حقوق صاحبان سهام}}{\text{کل دارایی‌ها}}$ ، ۱۴- کل بدهی‌ها به حقوق صاحبان سهام = $\frac{\text{کل بدهی‌ها}}{\text{حقوق صاحبان سهام}}$ ، ۱۵- گردش حساب‌های دریافتی = $\frac{\text{فروش}}{\text{حساب‌های دریافتی}}$ ، ۱۶- حاشیه عملیاتی = $\frac{\text{سود عملیاتی}}{\text{فروش}}$ ، ۱۷- گردش موجودی کالا = $\frac{\text{فروش}}{\text{موجودی کالا}}$ ، ۱۸- وجه نقد به کل دارایی‌ها = $\frac{\text{وجه نقد}}{\text{کل دارایی‌ها}}$ ، ۱۹- هزینه بهره به فروش = $\frac{\text{هزینه بهره}}{\text{فروش}}$ ، ۲۰- وجه نقد به بدهی‌های جاری = $\frac{\text{وجه نقد}}{\text{بدهی‌های جاری}}$ ، ۲۱- گردش دارایی‌های جاری = $\frac{\text{فروش}}{\text{دارایی‌های جاری}}$ ، ۲۲- گردش دارایی‌های ثابت = $\frac{\text{فروش}}{\text{دارایی‌های ثابت}}$ ، ۲۳- حاشیه سود ناخالص = $\frac{\text{سود ناخالص}}{\text{فروش}}$ ، ۲۴- بدهی‌های جاری به کل دارایی‌ها = $\frac{\text{بدهی‌های جاری}}{\text{کل دارایی‌ها}}$ ، ۲۵- سود هر سهم = $\frac{\text{سود خالص}}{\text{تعداد سهام عادی}}$ ، ۲۶- اندازه شرکت (لگاریتم کل دارایی‌ها)، ۲۷- گردش سرمایه در گردش = $\frac{\text{فروش}}{\text{سرمایه در گردش}}$ ، ۲۸- رشد کل دارایی‌ها، ۲۹- بدهی بلندمدت به کل دارایی‌ها = $\frac{\text{بدهی بلندمدت}}{\text{کل دارایی‌ها}}$	

$30 - \text{دارایی‌های سریع به کل دارایی‌ها} = \frac{\text{دارایی‌های سریع}}{\text{کل دارایی‌ها}}, 31 - \text{گردش حساب‌های پرداختی} =$
$\frac{\text{بهای تمام شده کالای فروش رفته}}{\text{حساب‌های پرداختی تجاری}}, 32 - \text{دارایی‌های سریع به فروش} = \frac{\text{دارایی‌های سریع}}{\text{فروش}}, 33 - \text{بدهی‌های}$
$\text{جاری به کل بدهی‌ها} = \frac{\text{بدهی‌های جاری}}{\text{کل بدهی‌ها}}, 34 - \text{فروش به موجودی کالا} = \frac{\text{فروش}}{\text{موجودی کالا}}, 35 - \text{دوره}$
$\text{حساب‌های دریافتی} = \frac{\text{حساب‌های دریافتی} \times 360}{\text{فروش}}, 36 - \text{سود خالص به دارایی‌های جاری} =$
$\frac{\text{سود خالص}}{\text{دارایی‌های جاری}}, 37 - \text{بدهی‌های بلندمدت به حقوق صاحبان سهام} = \frac{\text{بدهی‌های بلندمدت}}{\text{حقوق صاحبان سهام}}, 38 - \text{وجه}$
$\text{نقد به فروش} = \frac{\text{وجه نقد}}{\text{فروش}}, 39 - \text{رشد فروش}, 40 - \text{دارایی‌های ثابت به کل دارایی‌ها} = \frac{\text{دارایی‌های ثابت}}{\text{کل دارایی‌ها}},$
$41 - \text{گردش حقوق صاحبان سهام} = \frac{\text{فروش}}{\text{حقوق صاحبان سهام}}, 42 - \text{رشد سود خالص (در صورت وجود)}$
$\text{سود خالص در سال گذشته}: \frac{\text{سود (زیان) خالص سال جاری} - \text{سود خالص سال گذشته}}{\text{سود خالص سال گذشته}}, \text{در صورت وجود}$
$\text{زیان خالص در سال گذشته}: \frac{\text{سود (زیان) خالص سال جاری} + \text{زیان خالص سال گذشته} }{ \text{زیان خالص سال گذشته} }, 43 - \text{حقوق}$
$\text{صاحبان سهام به دارایی‌های ثابت} = \frac{\text{حقوق صاحبان سهام}}{\text{دارایی‌های ثابت}}, 44 - \text{بازده دارایی‌های ثابت} =$
$\frac{\text{سود خالص}}{\text{دارایی‌های ثابت}}, 45 - \text{سرمایه در گردش به حقوق صاحبان سهام} = \frac{\text{سرمایه در گردش}}{\text{حقوق صاحبان سهام}}, 46 - \text{سود}$
$\text{خالص به کل بدهی‌ها} = \frac{\text{سود خالص}}{\text{کل بدهی‌ها}}, 47 - \text{بدهی‌ها به خالص دارایی‌های مشهود} =$
$\frac{\text{بدهی‌ها}}{\text{خالص دارایی‌ها}}, 48 - \text{خالص دارایی‌ها به ازای هر سهم} = \frac{\text{خالص دارایی‌ها}}{\text{تعداد سهام}}$
$49 - \text{بدهی‌های جاری به حقوق صاحبان سهام} = \frac{\text{بدهی‌های جاری}}{\text{حقوق صاحبان سهام}}, 50 - \text{بهای تمام شده کالای}$
$\text{فروش رفته به فروش} = \frac{\text{بهای تمام شده کالای فروش رفته}}{\text{فروش}}, 51 - \text{اندازه شرکت (لگاریتم فروش)}, 52 -$
$\text{حاشیه سود عملیاتی قبل از استهلاک} = \frac{\text{سود عملیاتی قبل از استهلاک}}{\text{فروش}}, 53 - \text{بازده ناخالص کل}$
$\text{دارایی‌ها} = \frac{\text{سود ناخالص}}{\text{کل دارایی‌ها}}, 54 - \text{فروش به ازای کارکنان} = \frac{\text{فروش}}{\text{کل تعداد کارکنان}}, 55 - \text{تغییر در سود خالص}$
$= \frac{(NI_t - NI_{t-1})}{(NI_t + NI_{t-1})}, 56 - \text{موجودی کالا به کل دارایی‌ها} = \frac{\text{موجودی کالا}}{\text{کل دارایی‌ها}}, 57 - \text{رشد دارایی‌های}$
$\text{ثابت}, 58 - \text{تعدیلات سنواتی به کل دارایی‌ها} =$

$\frac{\text{تعدیلات سنواتی کل دارایی‌ها}}{\text{تعدیلات سنواتی هزینه مالیات}} = \frac{\text{تعدیلات سنواتی به فروش}}{\text{تعدیلات سنواتی فروش}} = ۵۹ -$ $\frac{\text{تعدیلات سنواتی به سود قبل از مالیات}}{\text{تعدیلات سنواتی هزینه مالیات}} = ۶۰ -$	
$\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۱ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۲ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۳ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۴ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۵ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۶ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۷ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۸ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۶۹ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۷۰ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۷۱ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۷۲ -$ $\frac{\text{جریان نقدی عملیاتی کل دارایی‌ها}}{\text{جریان نقدی عملیاتی کل بدهی‌ها}} = ۷۳ -$	نسبت‌های جریان نقدی
$\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری بدهی‌ها}} = ۷۴ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۷۵ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۷۶ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۷۷ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۷۸ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۷۹ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۰ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۱ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۲ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۳ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۴ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۵ -$ $\frac{\text{ارزش بازار سهام}}{\text{ارزش دفتری سهام}} = ۸۶ -$	متغیرهای بازار سهام و نقدشوندگی

<p>قدرمطلق بازده سهام t در روز t، 87- سهام شناور آزاد (حاصل ضرب ضریب شناوری در مجموع سهام سهامداران غیرمدیریتی).</p>	
<p>۸۸- تمرکز مالکیت ($HHI = \sum_{i=1}^{N_j} (SHARE_{i,j})^2$) که در این رابطه، $SHARE_{i,j}$: درصد سهام تحت تملک سهامدار i در شرکت j، 89- سهامداران عمده (مجموع درصد سهام تحت تملک سهامداران دارای مالکیت بیشتر از 5%)، 90- نخستین سهامدار بزرگ (درصد سهام تحت تملک نخستین سهامدار بزرگ دارای مالکیت بیشتر از 5%)، 91- دومین سهامدار بزرگ (درصد سهام تحت تملک دومین سهامدار بزرگ دارای مالکیت بیشتر از 5%)، 92- مالکیت مدیریتی (مجموع درصد سهام تحت تملک اعضای هیأت مدیره و مدیر عامل دارای مالکیت بیشتر از 5%)، 93- مالکیت خانوادگی (مجموع درصد سهام تحت تملک اعضای خانواده‌های دارای مالکیت بیشتر از 5%)، 94- مالکیت نهادی (مجموع درصد سهام تحت تملک سهامداران نهادی (بانک‌ها، شرکت‌های بیمه، صندوق‌های بازنشستگی، شرکت‌های سرمایه‌گذاری و سایر مؤسسه‌هایی که به خرید و فروش حجم بالایی از اوراق بهادار می‌پردازند) دارای مالکیت بیشتر از 5%)، 95- اندازه مؤسسه حسابرسی (بزرگ (سازمان حسابرسی) و کوچک (سایر مؤسسه‌های حسابرسی))، 96- نوع گزارش حسابرسی (مقبول، مشروط، عدم اظهارنظر، مردود)، 97- تغییر حسابرسی ((عدم) تغییر حسابرسی شرکت نسبت به سال گذشته)، 98- اندازه هیأت مدیره (تعداد اعضای هیأت مدیره)، 99- دوگانگی مدیر عامل (بکارگیری همزمان یک مدیر به عنوان مدیر عامل و رئیس هیأت مدیره)، 100- استقلال رئیس هیأت مدیره (مؤلف نبودن رئیس هیأت مدیره)، 101- استقلال هیأت مدیره</p> $= \frac{\text{تعداد اعضای غیرمؤلف هیأت مدیره}}{\text{تعداد کل اعضای هیأت مدیره}}$	<p>متغیرهای حاکمیتی شرکتی</p>
<p>102- رشد تولید ناخالص داخلی به قیمت پایه، 103- رشد تولید ناخالص ملی به قیمت پایه-ثابت، 104- رشد صادرات کالاها و خدمات-ثابت، 105- رشد واردات کالاها و خدمات-ثابت، 106- شاخص کل قیمت بورس، 107- رشد نقدینگی (پول + شبه پول)، 108- نرخ بیکاری، 109- نرخ بازار دلار، 110- رشد سرمایه‌گذاری خارجی، 111- نرخ تورم، 112- رشد مخارج دولت (کل پرداخت‌ها)، 113- چرخه تجاری (رونق)، 114- رتبه صندوق بین‌المللی پول، 115- شاخص کل تولید کارگاه‌های بزرگ صنعتی، 116- شاخص بهای کالاها و خدمات مصرفی، 117- شاخص بهای تولیدکننده، 118- نرخ اوراق مشارکت، 119- نرخ وام - صنعت و معدن، 120- نرخ بهره واقعی.</p>	<p>متغیرهای اقتصاد کلان</p>

۱۲۱- صنعت (طبقه‌بندی شرکت‌ها در ۳۷ صنعت بر اساس آیسیک ^۱)، ۱۲۲- سود انباشته به سرمایه، ۱۲۳- عمر شرکت بر حسب مدت پذیرش در بورس، ۱۲۴- عمر شرکت بر حسب زمان تأسیس.	سایر متغیرها
--	--------------

متغیر وابسته: در مطالعات تجربی مندرج در متون، ورشکستگی و درماندگی مالی به جای یکدیگر به کار گرفته شده‌اند. استفاده از درماندگی مالی موجب انعطاف‌پذیری در مرحله پژوهش می‌شود. درماندگی مالی در مقایسه با ورشکستگی تعریف متداول‌تری است و برای افزایش اندازه نمونه به پژوهش محدودیت‌هایی در پژوهش و کاهش اندازه نمونه می‌شود. استفاده از درماندگی مالی نه تنها در عمل بلکه در تئوری نیز دارای برتری است؛ زیرا همه شرکت‌های درمانده مالی ورشکسته نمی‌شوند. ورشکستگی آخرین گزینه برای شرکت‌هایی است که نمی‌توانند مشکلات مالی خودشان را حل کنند. به گونه خلاصه، استفاده از ورشکستگی تنها موجب محدود شدن به یک جنبه از درماندگی مالی می‌شود [133]. از این رو، در این پژوهش، برای تشریح وضعیت شرکت‌های روبه رو با دشواری مالی از عبارت درماندگی مالی استفاده می‌شود. افزون بر این، برای تشخیص شرکت‌های درمانده مالی نیز درماندگی مالی بر اساس ماده ۱۴۱ قانون تجارت ایران تعریف می‌شود [133]:

اگر بر اثر زیان‌های وارده حداقل نصف سرمایه شرکت از میان برود، هیأت مدیره مکلف است بلافاصله مجمع عمومی فوق‌العاده صاحبان سهام را دعوت نماید تا موضوع انحلال یا بقاء شرکت مورد شور و رأی واقع شود.

جامعه آماری، روش نمونه‌گیری و حجم نمونه

برای اجتناب از انتقادهای وارد بر جانبداری نمونه جفت تطبیقی، در این پژوهش از نمونه‌ای استفاده خواهد شد که نشان‌دهنده نرخ درماندگی واقعی است. از این رو، جامعه آماری این پژوهش، داده‌های ۴۲۱ شرکت پذیرفته شده در فرابورس و بورس اوراق بهادار تهران طی دوره ۱۶ ساله از سال ۱۳۹۵-۱۳۸۰ در تمامی صنایع به جز صنعت مالی است. این روش نمونه‌گیری، امکان ارزیابی تأثیر اندازه شرکت و صنعت بر احتمال درماندگی مالی را فراهم می‌کند. بر این اساس شرکت‌های خدمات مالی نظیر بانک‌ها، بیمه‌ها، کارگزاری‌ها، لیزینگ‌ها و غیره در تجزیه و تحلیل گنجانده نشده‌اند [24، 11، 3131، 21، 28، 1328]؛ زیرا شرکت‌های عضو این صنایع از دیدگاه ساختاری متفاوت هستند، محیط ورشکستگی متفاوتی دارند و در برخی موارد، بدست آوردن داده‌های مربوطه آن‌ها دشوار است [1328]. فهرست اسامی صنایع مورد بررسی در این پژوهش، در جدول ۲ نمایش داده شده است.

جدول ۲: صنایع مورد بررسی

ردیف	صنعت	ردیف	صنعت
۱	ایزار پزشکی، اپتیکی و اندازه‌گیری	۲۰	سایر محصولات کانی غیرفلزی
۲	استخراج زغال سنگ	۲۱	سیمان، آهک و گچ

¹ International Standard Industrial Classification (ISIC)

ردیف	صنعت	ردیف	صنعت
۳	استخراج سایر معادن	۲۲	عرضه برق، گاز، بخار و آب گرم
۴	استخراج کانه‌های فلزی	۲۳	فرآورده‌های نفتی، کک و سوخت هسته‌ای
۵	استخراج نفت گاز و خدمات جنبی به جز اکتشاف	۲۴	فلزات اساسی
۶	انبوه‌سازی، املاک و مستغلات	۲۵	قند و شکر
۷	انتشار، چاپ و تکثیر	۲۶	کاشی و سرامیک
۸	پیمانکاری صنعتی	۲۷	لاستیک و پلاستیک
۹	تجارت عمده‌فروشی به جز وسایل نقلیه موتور	۲۸	ماشین‌آلات و تجهیزات
۱۰	تجارت عمده و خرده‌فروشی وسائط نقلیه موتور	۲۹	ماشین‌آلات و دستگاه‌های برقی
۱۱	حمل و نقل، انبارداری و ارتباطات	۳۰	محصولات چوبی
۱۲	خدمات فنی و مهندسی	۳۱	محصولات شیمیایی
۱۳	خرده‌فروشی، به استثنای وسایل نقلیه موتوری	۳۲	محصولات غذایی و آشامیدنی به جز قند و شکر
۱۴	خودرو و ساخت قطعات	۳۳	محصولات کاغذی
۱۵	رایانه و فعالیت‌های وابسته به آن	۳۴	مخابرات
۱۶	زراعت و خدمات وابسته	۳۵	منسوجات
۱۷	ساخت دستگاه‌ها و وسایل ارتباطی	۳۶	مواد و محصولات دارویی
۱۸	ساخت محصولات فلزی	۳۷	هتل و رستوران
۱۹	سایر تجهیزات حمل و نقل		

ابزار گردآوری داده‌ها

در این پژوهش، برای نگارش و جمع‌آوری اطلاعات مورد نیاز بخش مبانی نظری و پیشینه پژوهش از مجله‌ها و کتاب‌های تخصصی لاتین، و برای گردآوری سایر داده‌های کمی و اطلاعات مورد نیاز از سایت کدال، سایت بورس اوراق بهادار تهران، بانک‌های اطلاعاتی سازمان بورس اوراق بهادار تهران، سایت بانک مرکزی جمهوری اسلامی ایران، سایت صندوق بین‌المللی پول، گزارش کلان اقتصادی پژوهشکده پولی و بانکی بانک مرکزی جمهوری اسلامی ایران، نرم افزار ره آورد نوین، نرم افزار تدبیر پرداز، و نرم افزار تی اس ای کلاینت^۱ استفاده خواهد شد. بنابراین، اطلاعات مورد نیاز بخش مبانی نظری و پیشینه پژوهش، داده ثانویه از نوع مستندات رسمی و داده‌های مورد نیاز بخش مدل‌سازی، داده ثانویه از نوع داده‌های پژوهشی آرشیوی است [1318].

^۱ TSEClient

روش تجزیه و تحلیل داده‌ها

پیش پردازش داده‌ها

تکنیک‌های پیش پردازش مورد استفاده در این پژوهش عبارت‌اند از:

پاک‌سازی داده: در این فرآیند، نخست رکوردهای تکراری حذف گردیدند اما به علت حجم اندک داده‌های در دسترس و زیاد بودن رکوردهای حاوی مقادیر گمشده، رکوردهای حاوی مقادیر گمشده حذف نشدند و بر حسب مورد به صورت دستی از طریق گزارش‌های مالی مندرج در سایت کدال و سایت بورس اوراق بهادار تهران، میانگین مقدار آن متغیر در سال قبل و بعد همان شرکت، یا مقدار آن متغیر در سال بعد همان شرکت تکمیل شده‌اند.

ترکیب داده: در این فرآیند، چندین منبع داده شامل ترازنامه، صورت سود و زیان، صورت جریان وجوه نقد، حاکمیت شرکتی، داده‌های بازار سهام و نقدشوندگی بر اساس دو شناسه نام (نماد) شرکت و سال و متغیرهای کلان اقتصادی بر اساس شناسه سال ادغام شده‌اند.

استخراج ویژگی‌های اولیه: ویژگی‌های استخراجی اولیه در جدول ۱، ارائه شده‌اند. با توجه به اینکه از یک سو، برای محاسبه برخی نسبت‌های مالی به اطلاعات سال گذشته نیاز است و از سوی دیگر، برای پیش‌بینی درماندگی مالی نیز از ویژگی‌های یک سال قبل استفاده می‌شود، نمونه انتخابی شامل ۳۶۷۰ سال-شرکت و ۱۲۳ ویژگی است.

وان هات انکدینگ: با توجه به اینکه ویژگی صنعت، یک متغیر طبقه‌ای اسمی هست، بر روی این ویژگی، وان هات انکدینگ اجرا شده است. پس از اجرای وان هات انکدینگ، به مجموعه داده، ۳۷ ستون جدید با مقادیر صرفاً ۰-۱ افزوده شده است. اما از آنجا که ویژگی نوع اظهار نظر حسابر، متغیر طبقه‌ای ترتیبی است، با استفاده از اعداد ۱ تا ۴ به مقادیر عددی تبدیل شده است.

ساخت ویژگی: ویژگی جدیدی تحت عنوان نسبت سود (زیان) انباشته به سرمایه سال قبل برای پیش‌بینی درماندگی مالی در سال بعد ساخته شده است.

سپس، داده آماده شده با استفاده تفکیک ۷۰ به ۳۰ به ترتیب به مجموعه داده آموزشی و آزمایشی تفکیک می‌شود [1329]. طی فرآیند آموزش، برای ارزیابی ترکیبات مختلف آبرپارامترهای الگوریتم یادگیری از اعتبارسنجی متقابل k لایه طبقه‌ای (با $k = 5$) استفاده خواهد شد.

مقیاس‌دهی ویژگی‌ها: در این پژوهش، برای مقیاس‌دهی ویژگی‌ها از استانداردسازی استفاده شده است. رویه استانداردسازی را می‌توان با معادله زیر بیان کرد که در این رابطه، μ_x میانگین نمونه یک ستون ویژگی خاص و σ_x انحراف معیار مربوط به آن است [1332].

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \quad (1)$$

لازم به ذکر است که استانداردسازی صرفاً بر روی داده آموزشی انجام و پارامترهای آن بر روی داده آزمایشی اعمال می‌شود.

انتخاب ویژگی: انتخاب زیرمجموعه ویژگی به معنای یافتن زیرمجموعه ویژگی I به گونه‌ای است که ریسک مورد انتظار طبقه کننده در صورت آموزش با I حداقل شود. در این پژوهش، انتخاب ویژگی به روش تعبیه‌ای انجام شده است [1330].

آموزش الگوریتم یادگیری

الگوریتم یادگیری: XGBoost. مخفف **Extreme Gradient Boosting**، سیستم یادگیری مبتنی بر بوستینگ گرادینت فریدمن (۲۰۰۱) است اما با برخی بهبودهایی که به ارتقا عملکرد کمک می‌کند. **XGBoost** از مدل دسته جمعی درختی استفاده می‌کند که مجموعه‌ای از درخت‌های طبقه‌بندی و رگرسیون (**CART**) است. این نوع بوستینگ، که از درخت‌ها به عنوان یادگیرنده‌های پایه استفاده می‌کند، را بوستینگ درختی می‌نامند. از آنجا که یک درخت برای دستیابی به نتایج خوب ممکن است کافی نباشد، چندین **CARTs** با هم می‌توانند مورد استفاده قرار گیرند و پیش‌بینی نهایی، مجموع هر یک از نمرات **CART** است. این مدل را می‌توان بدین گونه نوشت:

$$\hat{y}_i = \phi(X_i) = \sum_{k=1}^K f_k(X_i), f_k \in F \quad (2)$$

که در این رابطه، f تابعی در فضای تابعی F است به گونه‌ای که $F = \{f(x) = w_{q(x)}\} (q: R^m \rightarrow T, w \in R^T)$ مجموعه تمامی **CARTs** ممکن است که در این رابطه، q نشان‌دهنده ساختار هر درخت است که یک نمونه را به شاخص برگ مربوطه نگاشت می‌کند، T تعداد برگ‌ها در درخت است، w وزن برگ است و K نشان‌دهنده تعداد درخت‌ها است. تابع هدف مورد بهینه‌سازی به صورت تابع مندرج در معادله زیر در می‌آید و با افزودن f_t به شیوه تجمعی آموزش می‌بیند که به حداقل‌سازی هدف کمک می‌کند. در این رابطه، $\hat{y}_i^{(t-1)}$ نشان‌دهنده پیش‌بینی نمونه i در تکرار $t-1$ است، $l(y_i, \hat{y}_i^{(t-1)})$ تابع زیان آموزشی است و Ω عبارت تنظیم است [1327].

$$\mathcal{L}^{(t)} = \sum_i^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (3)$$

عبارت تنظیم با استفاده از معادله زیر محاسبه و برای کنترل واریانس برازش استفاده می‌شود تا انعطاف‌پذیری کار یادگیری کنترل شود و مدل‌هایی بدست آید که بهتر قابل تعمیم به داده دیده نشده است. کنترل پیچیدگی مدل به منظور اجتناب از برازش بیشتر از اندازه داده آموزشی سودمند است [132727].

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (4)$$

در حالی که در مدل گرادینت بوستینگ اولیه، درخت‌ها در قالب سری ساخته می‌شوند، این **XGBoost** کار را به صورت موازی و مشابه با روش جنگل تصادفی انجام می‌دهد که درخت‌ها به صورت موازی با

یکدیگر رشد می‌کنند یعنی، **XGBoost** طی مرحله آموزش با بکارگیری تمامی مراکز **CPU** کامپیوتر، خودش درخت را به صورت موازی می‌سازد که به افزایش سرعت محاسبات منجر می‌شود [1327].

یادگیری نامتوازن: در این پژوهش، با توجه به اینکه مجموعه داده درماندگی مالی، نامتوازن است، از تکنیک **SMOTE**^۱ استفاده شده است. این تکنیک نمونه‌هایی تصنعی از کلاس اقلیت بر اساس فاصله آن‌ها با یکدیگر در فضای ویژگی ایجاد می‌کند. هر نمونه اقلیت با یک خط به یکی یا تمامی **k** همسایه‌های نزدیک مرتبط می‌شود. تعداد نزدیک‌ترین همسایه‌های انتخابی به میزان نمونه‌گیری بیشتر از اندازه مورد نیاز بستگی دارد. از آنجا که **SMOTE** به صورت مستقیم نمونه‌های اقلیت را تکرار نمی‌کند بلکه نمونه‌های اقلیت جدیدی را تولید می‌کند، برخلاف نمونه‌گیری بیشتر از اندازه تصادفی، از برآزش بیشتر از اندازه اجتناب می‌کند [1335]. افزون بر این، نمره F_1 نیز به عنوان معیار ارزیابی عملکرد انتخاب می‌شود؛ زیرا نمره F_1 هنگام بررسی مسأله کلاس نامتوازن مناسب‌تر از **AUC** است [151315].

بهینه‌سازی آبرپارامتر

آبرپارامترها، پارامترهای مختص مدل هستند که برای بهینه‌سازی مدل مورد استفاده قرار می‌گیرند [1310]. بکارگیری اعتبارسنجی متقابل **k** لایه همراه با جستجوی شبکه‌ای، رویکرد مفیدی برای میزان‌سازی دقیق عملکرد مدل یادگیری ماشین از طریق تغییر مقادیر آبرپارامترهای آن هست [1332]. اما بهینه‌سازی آبرپارامترها با استفاده از جستجوی شبکه‌ای به روش اعتبارسنجی متقابل، فرآیندی مستلزم صرف منابع و زمان زیاد است. از این رو، در این پژوهش، فضای جستجوی آبرپارامترها و همچنین، تعداد تکرارهای اعتبارسنجی متقابل محدود شده است [1329]. پس از آموزش **XGBoost** با مقادیر مختلف آبرپارامترها و ارزیابی عملکرد آن، آبرپارامترهای مندرج در جدول ۳، به عنوان آبرپارامترهای بهینه مدل انتخاب شده‌اند که به مدل دارای بهترین عملکرد از نظر توان پیش‌بینی منجر می‌شوند.

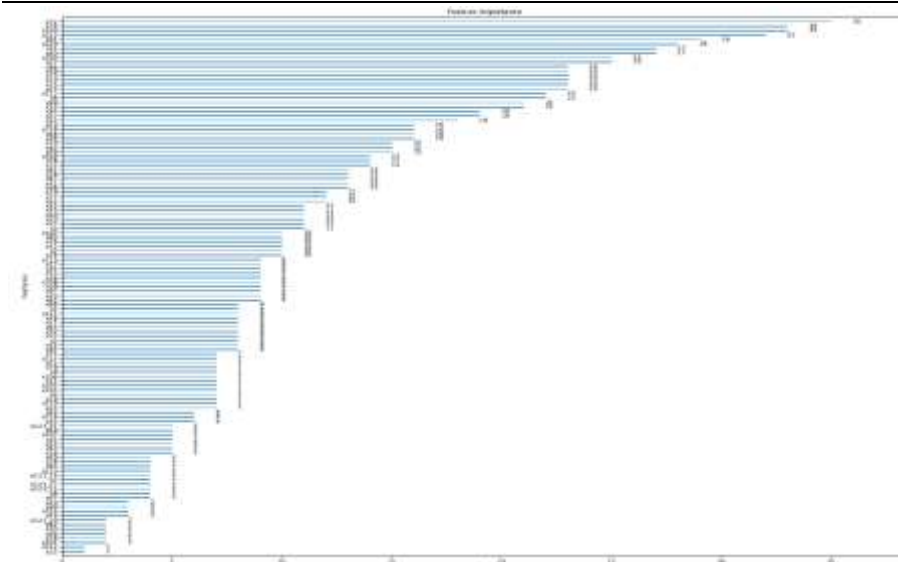
جدول ۳: آبرپارامترهای بهینه

مقادیر	آبرپارامترها	مقادیر	آبرپارامترها	مقادیر	آبرپارامترها
۰/۱۵	learning_rate	۱	subsample	۴	max_depth
۱	colsample_bylevel	۱۰۰	n_estimators	۱	colsample_bytree
۰	gamma	۰	reg_alpha	gbtree	booster
۴	scale_pos_weight	۱	min_child_weight	۲۰	reg_lambda
		۰/۵	base_score	۰	max_delta_step

پس از اجرای **XGBoost** با آبرپارامترهای بهینه بر روی داده‌های آموزشی، اهمیت ویژگی‌های نامربوط و اضافی در مدل، صفر می‌شود. میزان اهمیت ویژگی‌ها در شکل ۱ نمایش داده شده است. لازم به ذکر است که ویژگی‌های فاقد اهمیت در شکل نمایش داده نشده‌اند. بر اساس نتایج مندرج در شکل ۱، یک

¹ Synthetic Minority Over-sampling TEchnique (SMOTE)

سال قبل از درماندگی مالی، مهم‌ترین ویژگی‌های متمایزکننده شرکت‌های بالقوه درمانده و غیردرمانده در سال آتی عبارت‌اند از: مالیات نقدی پرداختی به سود قبل از مالیات، اندازه شرکت (لگاریتم ارزش بازار شرکت)، استقلال رئیس هیأت مدیره، سود انباشته به سرمایه، دومین سهامدار بزرگ، نرخ بهره واقعی، ریسک سیستماتیک، گردش حساب‌های پرداختی، گردش سرمایه در گردش، رشد تولید ناخالص ملی به قیمت پایه-ثابت، آمیهود، ارزش بازار سهام به ارزش دفتری سهام، سود تقسیمی هر سهم، تعدیلات سنواتی به فروش، بدهی‌ها به خالص دارایی‌های مشهود، سود انباشته به کل دارایی‌ها، حاشیه سود خالص، شاخص بهای تولیدکننده، دوگانگی مدیر عامل، کل بدهی‌ها به حقوق صاحبان سهام، مالکیت نهادی، رشد دارایی‌های ثابت، نسبت پوشش بهره، پوشش سرمایه‌گذاری مجدد، بازده ناخالص کل دارایی‌ها، سود خالص به دارایی‌های جاری، نرخ اوراق مشارکت، آمیوست، پوشش پرداخت سود سهام، بدهی‌های جاری به حقوق صاحبان سهام، ارزش بازار سهام به ارزش دفتری بدهی‌ها، تعدیلات سنواتی به کل دارایی‌ها، نرخ بازار دلار، اندازه هیأت مدیره، سهام شناور آزاد، تعداد معامله‌های سهام، گردش حقوق صاحبان سهام، رشد کل دارایی‌ها، بازدهی سهام، حاشیه سود عملیاتی قبل از استهلاک، هزینه بهره به فروش، هزینه مالیات به سود قبل از مالیات، تغییر در سود خالص، حقوق صاحبان سهام به دارایی‌های ثابت، بدهی‌های بلندمدت به حقوق صاحبان سهام، نسبت بدهی، حاشیه سود ناخالص، ارزش معامله‌های سهام، رشد سود خالص، اندازه شرکت (لگاریتم کل دارایی‌ها)، بازده دارایی‌ها، حاشیه عملیاتی و رشد تولید ناخالص داخلی به قیمت پایه است. همچنین، ویژگی‌های نوع گزارش حسابرسان، مالکیت خانوادگی، مالکیت مدیریتی، سهامداران عمده، بازده نقدی دارایی‌ها، بهای تمام شده کالای فروش رفته به فروش، حقوق صاحبان سهام به کل دارایی‌ها، عمر شرکت بر حسب مدت پذیرش در بورس، تمامی صنایع به جز خودرو و ساخت قطعات، سایر محصولات کانی غیرفلزی، قند و شکر، لاستیک و پلاستیک، محصولات شیمیایی و چرخه تجاری (رونق) در پیش‌بینی درماندگی مالی سال آتی فاقد اهمیت هستند.



شکل ۱. اهمیت ویژگی ها

سپس، مدل انتخابی، بر روی کل ۷۰٪ داده آموزشی تفکیک شده، آموزش داده می‌شود. آموزش مدل با استفاده از کل مجموعه داده آموزشی، به ماتریس درهم ریختگی مندرج در جدول ۶، منجر می‌شود.

جدول ۶: ماتریس درهم ریختگی مجموعه داده آموزشی

واقعی		وضعیت	
تعداد شرکت سالم	تعداد شرکت درمانده	پیش‌بینی شده	
۲۴۵۳	۰		
۲۱	۹۵	جمع	
۲۴۷۴	۹۵		

معیارهای ارزیابی عملکرد طبقه‌بندی کننده انتخابی بر روی کل مجموعه داده آموزشی در جدول ۷ نشان داده شده است.

جدول ۷: معیارهای ارزیابی مجموعه داده آموزشی

معیار	مقدار	معیار	مقدار	معیار	مقدار
صحت	٪۹۹	ضریب همبستگی متیوز	٪۹۰	معیار	٪۸۲
فراخوانی	٪۱۰۰	میانگین هندسی	٪۱۰۰	معیار F	٪۹۰
کاپا کوهن	٪۹۰	شاخص صحت متوازن	٪۸۳	خطای نوع ۱	٪۰

کاربرد XGBoost برای پیش‌بینی درماندگی مالی...

۲۱.

خطای نوع ۲	٪۰.۱	ناحیه زیر منحنی مشخصه عملیاتی دریافت کننده	٪۱۰۰
------------	------	--	------

ارزیابی مدل نهایی

ارزیابی مدل نهایی به معنای برآورد عملکرد مدل انتخابی بر روی داده جدید است. از این رو، عملکرد مدل انتخابی در قسمت قبل، بر روی ۳۰٪ داده آزمایشی کنار گذاشته شده، مورد آزمون قرار می‌گیرد. پس از ارائه مجموعه داده آزمایشی به مدل انتخابی، ماتریس درهم ریختگی مندرج در جدول ۸ بدست می‌آید.

جدول ۸: ماتریس درهم ریختگی مجموعه داده آزمایشی

وضعیت		واقعی	
		تعداد شرکت سالم	تعداد شرکت درمانده
پیش‌بینی شده	تعداد شرکت سالم	۱۰۱۹	۱۱
	تعداد شرکت درمانده	۴۲	۲۹
جمع		۱۰۶۱	۴۰

معیارهای ارزیابی حاصل از ماتریس درهم ریختگی مجموعه داده آزمایشی به شرح مندرج در جدول ۹ است.

جدول ۹: معیارهای ارزیابی مجموعه داده آزمایشی

معیار	مقدار	معیار	مقدار	معیار	مقدار
صحت	٪۹۵	ضریب همبستگی متیوز	٪۵۲	دقت	٪۴۱
فراخوانی	٪۷۳	میانگین هندسی	٪۸۳	معیار F	٪۵۲
کاپا کوهن	٪۵۰	شاخص صحت متوازن	٪۵۱	خطای نوع ۱	٪۲۸
خطای نوع ۲	٪۰.۴	ناحیه زیر منحنی مشخصه عملیاتی دریافت کننده	٪۹۶		

نتیجه‌گیری و ارایه پیشنهادها

درماندگی مالی، رویداد ویرانگری است که نه تنها منافع طیف وسیعی از ذینفعان را از بین می‌برد بلکه بر توسعه پایدار اقتصاد و جامعه نیز تأثیر می‌گذارد. از این رو، هدف این پژوهش، ارائه مدلی برای پیش‌بینی درماندگی مالی شرکت‌های پذیرفته شده در فرابورس و بورس اوراق بهادار تهران است. بر اساس نتایج این پژوهش، مدل XGBoost به نمره $F1$ ، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۹۰٪، ۹۰٪، ۱۰۰٪ و ۸۲٪ بر روی مجموعه آموزشی منجر می‌شود. افزون بر این، آزمون مدل پیشنهادی بر روی

مجموعه آزمایشی کنار گذاشته شده حاکی از این است که نمره **F1**، ضریب همبستگی متیوز، فراخوانی و دقتی به ترتیب برابر با ۵۲٪، ۷۲٪ و ۴۱٪ بر روی مجموعه آزمایشی بدست می‌آید. بنابراین، یافته‌های این پژوهش حاکی از این است که اطلاعات صورت‌های مالی، بازار سهام، حاکمیت شرکتی و متغیرهای اقتصاد کلان می‌تواند به عنوان ابزاری برای علامت‌دهی در ماندگی مالی مورد استفاده قرار گیرد. از سوی دیگر، پژوهش بیور و همکاران (۲۰۰۵) نیز بیانگر این است که محتوای اطلاعاتی داده‌های صورت‌های مالی برای هشدار در ماندگی مالی در طی زمان کاسته نشده است. از این رو، یافته‌های این پژوهش با یافته‌های آن‌ها سازگار است اگر چه این پژوهش‌ها در محیط متفاوت و با استفاده از تکنیک‌های تحلیلی متفاوتی در دوره‌های زمانی مختلفی انجام شده‌اند. همچنین، یافته‌های این پژوهش دارای مفاهیم ضمنی برای تحلیل‌گران و مدیران بخش دولتی و خصوصی است. با بکارگیری مدل پیش‌بینی در ماندگی شرکت‌ها، قانون‌گذاران می‌توانند به راحتی ثبات اقتصادی را نظارت و تضمین کنند، سرمایه‌گذاران می‌توانند تصمیم‌های بهتری در خصوص زمان خرید سهام جدید یا حتی فروش سهام بگیرند، بستانکاران نیز از ارزش اعتباری شرکت‌ها آگاهی می‌یابند و می‌توانند بهتر در خصوص اعطای اعتبار، شرایط و مدت آن قضاوت نمایند، مدیریت شرکت می‌تواند از دست دادن مدیران و کارگران کلیدی، عرضه‌کنندگان عمده، مشتریان بزرگ و همچنین، از بین رفتن اعتماد بستانکاران، سهامداران و سرمایه‌گذاران پیشگیری نماید.

در انجام این پژوهش محدودیت‌هایی به شرح زیر وجود داشت: از یک سو، پذیرش تعریف قانونی از در ماندگی مالی طبق ماده ۱۴۱ قانون تجارت، و از سوی دیگر، احتمال دست‌کاری حساب‌ها، ابهاماتی را در خصوص تمایز دقیق شرکت‌های درمانده از غیردرمانده مطرح می‌کند. به بیان دقیق‌تر، نه تنها خود شرکت‌ها تمایل دارند که به منظور عدم شمول ماده ۱۴۱ قانون تجارت، صورت‌های مالی خود را دست‌کاری نمایند بلکه قانون‌گذار نیز انتقال مازاد تجدید ارزیابی شرکت‌ها به حساب افزایش سرمایه را برای شرکت‌های مشمول ماده ۱۴۱ قانون تجارت که به وسیله این تجدید ارزیابی از ماده ۱۴۱ قانون تجارت خارج شوند، را مجاز می‌داند هر چند که از لحاظ محتوایی و بنیادی، تغییر اساسی در ساختار مالی شرکت‌ها ایجاد نمی‌شود و صرفاً نوعی حساب آرایبی است. بدیهی است که عدم تفکیک دقیق شرکت‌های درمانده از غیردرمانده تأثیر منفی بر عملکرد مدل می‌گذارد. افزون بر این، ماهیت نوپزی داده‌های حسابداری و مالی، دست‌کاری حساب‌ها، تجدید ارائه مکرر صورت‌های مالی و در نتیجه تعدیلات سنواتی زیاد، عدم رعایت یکنواختی و ثبات رویه در تهیه گزارش‌های مالی و همچنین، دانش و دقت ناکافی دست‌اندرکاران تهیه و ارائه صورت‌های مالی در طبقه‌بندی صحیح حساب‌ها در سرفصل‌های مربوطه موجب کاهش محتوای اطلاعاتی داده‌های پیش‌بینی‌کننده و در نتیجه، تخریب عملکرد مدل می‌شود. از این رو، پیشنهاد می‌شود که با افزودن داده‌های سال‌های آتی شرکت‌های فعلی و همچنین، شرکت‌های جدیدی که سهام آن‌ها در آینده در بورس اوراق بهادار تهران معامله خواهد شد، حجم داده‌های مورد استفاده را افزایش دهند تا بتوانند با بکارگیری الگوریتم‌های غیرخطی نظیر مدل‌های یادگیری عمیق، عملکرد مدل را بهبود دهند.

فهرست منابع

۱. احمدپور، احمد و حبیبه میرزایی اسرمی (۱۳۹۲). مقایسه مدل تحلیل تمایزی چندگانه با مدل شبکه‌های عصبی در پیش‌بینی ورشکستگی شرکت‌های پذیرفته شده در بازار بورس اوراق بهادار تهران. *فصلنامه تحقیقات حسابداری و حسابرسی*، سال پنجم، شماره ۱۸، ۲۲-۴.
۲. کاتبی، حسینقلی (۱۳۸۰). *حقوق تجارت*. چاپ هفتم، تهران: انتشارات گنج دانش
3. Aktan, S. (2011). Early warning system for bankruptcy: Bankruptcy prediction (Doctoral dissertation, Karlsruhe Institute of Technology, KIT). Retrieved from <https://d-nb.info/1019790032/34>.
4. Altman, E. I. (1984). "The success of business failure prediction models: An international survey", **Journal of Banking and Finance**, 8(2), 171-198.
5. Altman, E. I., & Narayanan, P. (1997). "An international survey of business failure classification models", **Financial Markets, Institutions & Instruments**, 6(2), 1-57.
6. Barboza, F., Kimura, H., & Altman, E. (2017). "Machine learning models and bankruptcy prediction", **Expert Systems with Applications**, 83, 405-417.
7. Beaver, W. H., Correia, M., & McNichols M. F. (2010). "Financial statement analysis and the prediction of financial distress", **Foundations and Trends in Accounting**, 5(2), 99-173.
8. Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). "Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy", **Review of Accounting Studies**, 10(1), 93-122.
9. Belli, G. (2009). Nonexperimental quantitative research. In S. D. Lapan & M. T. Quartaroli (Eds.), *Research essentials: An introduction to designs and practices*. (pp. 59-77). Jossey-Bass Publications.
10. Bonnes, K. (2017). Predicting mortgage demand using machine learning techniques (Master Thesis, University of Twente). Retrieved from https://essay.utwente.nl/73640/7/Bonnes_MA_EEMCS.pdf.
11. Chancharat, N. (2008). An empirical analysis of financially distressed Australian companies: The application of survival analysis (Doctoral dissertation, University of Wollongong). Retrieved from <https://ro.uow.edu.au/theses/401/>.
12. Charalambakis, E. C., & Garrett, I. (2018). "On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece", **Review of Quantitative Finance and Accounting**, 52(2), 467-491.
13. Cho, S., Kim, J., & Bae, J. K. (2009). "An integrative model with subject weight based on neural network learning for bankruptcy prediction", **Expert Systems with Applications**, 36(1), 403-410.
14. Etemadi, H., Anvary Rostamy, A. A., & Farajzadeh Dehkordi, H. (2009). "A genetic programming model for bankruptcy prediction: Empirical evidence from Iran", **Expert Systems with Applications**, 36(2), 3199-3207.
15. Fan, X. (2016). An adaptive and diversity-based ensemble method for binary classification (Master Thesis, Carleton University). Retrieved from

https://curve.carleton.ca/system/files/etd/e2e72ad2-2da5-47b8-ab6a-28810d5eb197/etd_pdf/81f038dc409086dafcc1cced7b46be03/fan-anadaptiveanddiversitybasedensemblemeth.pdf.

16. Jaikengkit, A. (2004). Corporate governance and financial distress: an empirical analysis - the case of Thai financial institutions (Doctoral dissertation, Case Western Reserve University). Retrieved from ProQuest Dissertations and Theses A&I. (AAT 3118137).
17. Jantadej, P. (2006). Using the combinations of cash flow components to predict financial distress (Doctoral dissertation, University of Nebraska). Retrieved from ProQuest Dissertations and Theses A&I. (AAT 3216429).
18. Johnson, R. B., & Christensen, L. (2014). *Educational research: quantitative, qualitative, and mixed approaches* (5th ed.). London: SAGE Publications, Inc.
19. Kothari, C. R. (2004). *Research methodology, methods and techniques*. New Delhi: New Age International (P) Ltd. Publishers.
20. Leano, H. J. (2004). Discriminant analysis, factor analysis and regression analysis to classify financially distressed firms and predict bankruptcy using financial ratios and macroeconomic predictors: Model application to selected M&A (Master Thesis, Lamar University). Retrieved from ProQuest Dissertations and Theses A&I (AAT 1426146).
21. Lee, K., Booth, D., & Alam, P. (2005). "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms", **Expert Systems with Applications**, 29(1), 1–16.
22. Lee, T. S., Yeh, Y. H., & Liu, R. T. (2003). *Can corporate governance variables enhance the prediction power of accounting-based financial distress prediction models?* (Center for Economic Institutions Working Paper Series, No. 2003-14). Retrieved from <https://core.ac.uk/download/pdf/7063883.pdf>.
23. Lensberg, T., Eilifsen, A., & McKee, T. E. (2006). "Bankruptcy theory development and classification via genetic programming", **European Journal of Operational Research**, 169(2), 677–697.
24. Li, M. Y. L., & Miu, P. (2010). "A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach", **Journal of Empirical Finance**, 17(4), 818–833.
25. Lin, R. H., Wang, Y. T., Wu, C. H., & Chuang, C. L. (2009). "Developing a business failure prediction model via RST, GRA and CBR", **Expert Systems with Applications**, 36(2), 1593–1600.
26. McKee, T. E., & Lensberg, T. (2002). "Genetic programming and rough sets: A hybrid approach to bankruptcy classification", **European Journal of Operational Research**, 138(2), 436–451.
27. Nobre, J., & Neves R. (2019). "Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets", **Expert Systems with Applications**, 125, 181–194.
28. Ohlson, J. A. (1980). "Financial ratios and the probabilistic prediction of bankruptcy", **Journal of Accounting Research**, 18(1), 109–131.

29. Oleksy, T. A. (2017). Machine learning methods for mood disorder decision support (Master Thesis, University of Bergen). Retrieved from <http://bora.uib.no/bitstream/handle/1956/16259/actigraphdataformoods.pdf?sequence=4&isAllowed=y>.
30. Pajkossy, K. (2013). Studying feature selection methods applied to classification tasks in natural language processing (Master Thesis, Eotvos Lorand University of Sciences). Retrieved from https://web.cs.elte.hu/blobs/diplomamunkak/msc_alkmat/2013/pajkossy_katalin.pdf
31. Pendharkar, P. C. (2005). "A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem", **Computers & Operations Research**, 32(10), 2561–2582.
32. Raschka, S. (2015). *Python machine learning*. Birmingham: Packt Publishing Ltd..
33. Schellenger, M., & Cross, J. N. (1994). "FASB 95, Cash flow and bankruptcy", **Journal of Economics and Finance**, 18(3), 261–274.
34. Suntraruk, Phassawan (2009). Predicting Financial Distress: Evidence from Thailand. <http://www.efmaefm.org/OEFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2009-milan/phd/phassawan.pdf>
35. Uguroglu, S. (2013). Robust Learning with Highly Skewed Category Distributions (Doctoral dissertation, Carnegie Mellon University). Retrieved from <https://pdfs.semanticscholar.org/c792/d83d78ff10b7a944d2c4e534c2c55bdf59b8.pdf>
36. Wu, W. W. (2010). "Beyond business failure prediction", **Expert Systems with Applications**, 37(3), 2371–2376.
37. Wu, Y., Gaunt, C., & Gray, S. (2010). "A comparison of alternative bankruptcy prediction models", **Journal of Contemporary Accounting & Economics**, 6(1), 34–45.
38. Zhou, L., Lai, K. K., & Yen, J. (2012). "Empirical models based on features ranking techniques for corporate financial distress prediction", **Computers and Mathematics with Applications**, 64(8), 2484–2496.



Application of XGBoost to Predict Financial Distress of the Listed Companies on Tehran Stock Exchange (TSE) and Iran Fara Bourse (IFB)

Shahla Ebrahimi

PhD Student in Accounting, Department of Accounting, Mazandaran University, Babolsar, Iran

Mohammad Namazi(PhD)¹

Distinguished Professor, Department of Accounting, Mazandaran University, Babolsar, Iran

(Received: 5 April 2020; Accepted: 27 June 2020)

The purpose of this article is to predict the potential financial distress of the listed companies on Tehran Stock Exchange (TSE) and Iran Fara Bourse (IFB). To do so, a wide range of features including accrual accounting variables, cash-based accounting variables, market-based variables, corporate governance mechanisms, and macroeconomic indicators have been identified to prospectively predict the financial distress in the companies.

The final sample includes 421 firms leading to 3,670 firm-year observations. The prepared data, was then split into a train and test data set using a 70/30 ratio.

In this research, various data pre-processing machine learning techniques i.e., Z-score standardization, one-hot encoding, stratified K-fold validation combined with feature engineering are applied to improve classifier performance. Stratified K-fold cross validation method, (with $k = 5$) was used for estimation of model prediction performance during training phase. During the training phase, hyper-parameter tuning of a model was carried out using a grid-search. Furthermore, SMOTE technique in conjunction with the proposed imbalance-oriented metric i.e., F1 score were used to overcome the extreme class imbalance issue.

Based on the experimental results, the tuned XGBoost model achieved a f1-score, MCC, recall and precision of respectively, 90%, 90%, 100% and 82% on the training set. Finally, the proposed model was tested on the hold-out test set which resulted in a f1-score, MCC, recall and precision of 52%, 52%, 73% and 41%, respectively. This information provides a powerful tool for predicting the financial distress of companies.

Keywords: Financial Distress Prediction, XGBoost, Machine Learning, Data Mining, Tehran Stock Exchange.

¹ mnamazi@rose.shirazu.ac.ir © (Corresponding Author)